

Metadata Records Creation Using ERC and DC Simple

Introduction

This document should demonstrate the author's increasing understanding of two metadata schema, the Electronic Resource Citation (ERC) scheme developed by John Kunze, and the Dublin Core (DC) Simple scheme developed by the Dublin Core Metadata Initiative (DCMI) community. Specifically, this document will show the author's learning of each scheme's elements and syntax, and discovery and working solutions to issues encountered during input of data for both schema. Ten total records—consisting of five records each of ERC and DC—are presented in this report, followed by discussion of author's observations leading up to and during input.

Electronic Resource Citation Records

Record 1

erc:

who: National Information Standards Organization

what: Understanding Metadata

when: 2004

where: <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>

erc-about:

what: Metadata | Technical standards

Record 2

erc:

who: Le Boeuf, Patrick

what: Brave New FRBR World

when: :unkn

where: http://www.ddb.de/news/pdf/papers_leboeuf.pdf

erc-about:

what: FRBR

****Note:** date not found on object at hand. This is addressed in the text of this document on pages 8-9.

Record 3

erc:

who: International Digital Object Identifier Foundation

what: Digital Object Identifier System

when: 20041123

where: <http://www.doi.org/index.html>

Record 4

erc:

who: Research Libraries Group

what: Trusted Digital Repositories: Attributes and Responsibilities

when: 2002

where: <http://www.rlg.org/longterm/repositories.pdf>

Record 5

erc:

who: Miller, Eric

what: An Introduction to the Resource Description Framework

when: 199805

where: <http://www.dlib.org/dlib/may98/miller/05miller.html>

Dublin Core Records

Record 1

<META NAME="DC.Title" LANG="en" CONTENT="Understanding Metadata">

<META NAME="DC.Creator" LANG="en" CONTENT="National Standards Information Organization">

<META NAME="DC.Subject" LANG="en" CONTENT="metadata">

<META NAME="DC.Subject" LANG="en" CONTENT="technical standards">

<META NAME="DC.Description" LANG="en" CONTENT="An exploration of metadata (structured data about data), including an expanded definition, common schema and element sets, protocols for interoperability and exchange, and predicted future uses and directions, as well as a glossary and resource list.">

<META NAME="DC.Publisher" LANG="en" CONTENT="National Standards Information Organization">

<META NAME="DC.Date" LANG="en" CONTENT="2004">

<META NAME="DC.Type" LANG="en" CONTENT="text">

<META NAME="DC.Format" LANG="en" CONTENT="application/pdf">

<META NAME="DC.Identifier" LANG="en"
CONTENT="http://www.niso.org/standards/resources/UnderstandingMetadata.pdf">

<META NAME="DC.Language" LANG="en" CONTENT="en">

Record 2

<META NAME="DC.Title" LANG="en" CONTENT="Brave New FRBR World">

<META NAME="DC.Creator" LANG="en" CONTENT="Le Boeuf, Patrick">

<META NAME="DC.Subject" LANG="en" CONTENT="FRBR">

<META NAME="DC.Subject" LANG="en" CONTENT="Functional requirements
for bibliographic standards">

<META NAME="DC.Description" LANG="en" CONTENT="White paper about FRBR, a model
for the functional requirements of bibliographic records.">

<META NAME="DC.Publisher" LANG="en" CONTENT="Die Deutsche Bibliothek">

<META NAME="DC.Type" LANG="en" CONTENT="text">

<META NAME="DC.Format" LANG="en" CONTENT="application/pdf">

<META NAME="DC.Identifier" LANG="en"
CONTENT="http://www.ddb.de/news/pdf/papers_leboeuf.pdf">

<META NAME="DC.Language" LANG="en" CONTENT="en">

Record 3

<META NAME="DC.Title" LANG="en" CONTENT="Welcome to the Digital Object Identifier
System ">

<META NAME="DC.Title" LANG="en" CONTENT="The Digital Object Identifier System">

<META NAME="DC.Creator" LANG="en" CONTENT="International DOI Foundation">

<META NAME="DC.Subject" LANG="en" CONTENT="digital object identifiers">

<META NAME="DC.Description" LANG="en" CONTENT="The Digital Object Identifier (DOI)
website concerns a naming system for identifying content objects and their current
information in the digital environment.">

<META NAME="DC.Publisher" LANG="en" CONTENT="International DOI Foundation">

<META NAME="DC.Date" LANG="en" CONTENT="2004-11-23">

<META NAME="DC.Type" LANG="en" CONTENT="text">

<META NAME="DC.Format" LANG="en" CONTENT="text/html">

<META NAME="DC.Identifier" LANG="en" CONTENT="http://www.doi.org/index.html">

<META NAME="DC.Language" LANG="en" CONTENT="en">

Record 4

<META NAME="DC.Title" LANG="en" CONTENT="Trusted Digital Repositories: Attributes and Responsibilities">

<META NAME="DC.Creator" LANG="en" CONTENT="The Research Libraries Group, Incorporated">

<META NAME="DC.Subject" LANG="en" CONTENT="digital repositories">

<META NAME="DC.Subject" LANG="en" CONTENT="digital resources management">

<META NAME="DC.Subject" LANG="en" CONTENT="digital archives">

<META NAME="DC.Description" LANG="en" CONTENT="Guidelines for creating, certifying and managing a trusted digital repository or archive.">

<META NAME="DC.Publisher" LANG="en" CONTENT="The Research Libraries Group, Incorporated">

<META NAME="DC.Contributor" LANG="en" CONTENT="Online Computer Library Center">

<META NAME="DC.Date" LANG="en" CONTENT="2002-05">

<META NAME="DC.Type" LANG="en" CONTENT="text">

<META NAME="DC.Format" LANG="en" CONTENT="application/pdf">

<META NAME="DC.Identifier" LANG="en"
CONTENT="http://www.rlg.org/longterm/repositories.pdf">

<META NAME="DC.Language" LANG="en" CONTENT="en">

Record 5

<META NAME="DC.Title" LANG="en" CONTENT="An Introduction to the Resource Description Framework">

<META NAME="DC.Creator" LANG="en" CONTENT="Miller, Eric">

<META NAME="DC.Subject" LANG="en" CONTENT="Resource Description Framework">

```
<META NAME="DC.Subject" LANG="en" CONTENT="XML application">
```

```
<META NAME="DC.Subject" LANG="en" CONTENT="RDF">
```

```
<META NAME="DC.Description" LANG="en" CONTENT="May 1998 D-Lib Magazine article  
introducing reader to the Resource Description Framework (RDF), an infrastructure that  
enables the encoding, exchange and reuse of structured metadata.">
```

```
<META NAME="DC.Publisher" LANG="en" CONTENT="D-Lib Magazine">
```

```
<META NAME="DC.Date" LANG="en" CONTENT="1998-05">
```

```
<META NAME="DC.Type" LANG="en" CONTENT="text">
```

```
<META NAME="DC.Format" LANG="en" CONTENT="text/html">
```

```
<META NAME="DC.Identifier" LANG="en"  
CONTENT="http://www.dlib.org/dlib/may98/miller/05miller.html">
```

```
<META NAME="DC.Language" LANG="en" CONTENT="en">
```

Discussion

In order to begin the project that this document stems from, the author located and read several documents pertaining to the architecture, standards, elements and semantics of both the ERC and DC Simple metadata schema. Those documents are listed in the "Works Consulted" list at the end of this report. The author also located and downloaded the generic editing software "NoteTab Light 4.95" and two ClipBook Template (.clb) files to compose a metadata creation tool. This tool allowed the author to input data in the element order suggested by each scheme. Since Dublin Core demands no set order of elements, the order followed was the order in which the elements were ordered within the metadata tool. Because ERC does impose a specific order, at least in its anchoring story, that order was followed, an endeavor helped by the use of this specific metadata creation tool. This tool also provided the appropriate syntax for each scheme—a time-saving and error-proofing benefit for the data entry person; once the rules for input of data for each element were understood, the only thing left was to fill in the blanks, and uniformly so.

There were three discrete but related sets of problems presented through further investigation of proper element content creation: one set of problems was related to the appropriateness of data for each element, another set of problems originated in the location of that data, and the third set in uniform entry of those data. The first problem set, in this author's humble opinion, is caused in part by the subjective nature of descriptive metadata, and by difficulty in concept articulation and in the creation of an elegantly Spartan but logically inclusive set of elements to enable description of an astoundingly wide array of objects across many domains. The second problem set is of a more practical nature, e.g. "where on this thing is the data to fill in that blank?" This author's decision-making on the second set was based on the old idea of using the object at hand. The third problem centered on the establishment of data-entry rules. All three problem sets are faced by catalogers, and others tasked with metadata, on a constant basis. Discussion of all three sets follows; though multiple instances of each problem were encountered, for the purpose of brevity only one instance per each set is discussed in detail.

Navigating the minefield of rules for each element was by far the most time-consuming and treacherous part of this endeavor. For instance upon first encounter the elements for ERC seem stultifyingly simple: all one needs to know is "who", "what", "when" and "where". However, upon deeper investigation, fewer elements make for more complexity in determining the appropriate data, not less. This is especially true when multiple values can be used to fill in the blanks. "Which 'who' should I use" one wonders, "the first author's name, or all the authors' names, or their organization(s), or the document's publisher?" Fortunately, each basic element in ERC can be modified to fit within its so-called "anchoring story", a concept which itself is elusive and only vaguely described by its creator John Kunze, but is an asset because of its semantic flexibility. Similar problems are found within certain elements (but not all) of DC Simple. The interplay between the "Creator" and "Publisher" and "Contributor" elements can be confusing. The "Contributor" guidelines in the "Using Dublin Core" document by Diane Hillmann state,

The same general guidelines for using names of persons or organizations as Creators apply here. Contributor is the most general of the elements used for "agents" responsible for the resource, so should be used when primary responsibility is unknown or irrelevant. [emphasis mine]

One wonders why, since the "Contributor" element is stated here to be 'the most general of elements', it should be used at all if primary responsibility is unknown or irrelevant. Why use the element at all if the value is determined to be irrelevant? And what permutation of "agents responsible for the resource" is not covered by listing multiple creators? Since the "Creator" and "Contributor" elements have very similar guidelines, these are vexing questions indeed. Creation of content according to the guidelines for "Publisher" was more straightforward, when this information was listed separately on the document; since self-publication is so much more ubiquitous in the digital world than the atom-based "real" world, data to be entered into the "Creator" element and the "Publisher" element are often the same, but also are often obscured on the document even if different agents are responsible.

Similarly, headaches were caused by lack of explicit information on the documents themselves. One document in particular for which metadata was created was "Brave New FRBR World" by Patrick Le Bœuf. It had no date of publication or creation or modification stated explicitly on its pages. Since it was presented in this instance in Portable Document Format (PDF), it was not possible to obtain the date (as is sometimes possible on html-coded pages) by viewing the page source code. The year of publication could be surmised as either 2003 or 2004 or 2005, since those were the intervening years between the latest event referenced within the text of his document and now, but there is no way to narrow the actual date down completely and also no way to account for such *partially known* values within either ERC or DC Simple. Additionally, a reference was found to the Le Bœuf document within another document (specifically within Dr. William Moen's online class assignment notes), but since this reference was not obtained from the Le Bœuf document (the document in hand) then the referred date was disregarded. Consequently, the ERC record of the Le Bœuf document shows the date as the value equivalent in meaning to "unknown",

and the DC record for the same document omits the date element entirely. It follows that this might cause a problem with permanence of the record were it in actual use as metadata, since no date is associated with the document. One would think that another date, the date of metadata record creation, could then be used in lieu of a document-based date, but this remains unclear presently.

A third, easily-solved, problem was encountered upon an attempt to ensure the uniformity and quality of the data entered: entry rules must be followed, particularly if the resulting records are expected to be uniform and therefore consistently sorted by a machine. For example, in Hillmann's "Using Dublin Core", the convention for the DC element "Creator" is stated:

Creators should be listed separately, preferably in the same order that they appear in the publication. Personal names should be listed surname or family name first, followed by forename or given name. When in doubt, give the name as it appears, and do not invert.

So accordingly, personal names were entered last name first into the "Creator" element in each record where appropriate. The rules do not explicitly state that a comma should be used between last and first names, so a decision was made to place a comma between to make it clear that reverse name order is being imposed. It is thought that most people, when encountering a name listed in forward (i.e. first then last) order, will realize that the name is in such order because there is *no* punctuation, and that conversely, placement of a comma between names denotes reverse order. The exact same use of a comma could also denote a list of multiple authors' names, but this is very uncommon, especially since such use in this case would show either only the first names or only the last names of each author.

Summary and Conclusion

So in learning to encode metadata according to ERC and DC Simple standards, three major issues were encountered, issues which can be reduced to object description, information location, and input rules—all of which are old issues that have been facing catalogers for decades. The issues are not new; however, they necessarily must be

approached within a new and markedly different context, the context of digital objects versus the context of atom-based materials that were cataloged in the past. Because of such phenomena as the ubiquity of digital self-publication (causing problems with what amount to statements of responsibility), and the lack of information and/or the presence of competing bits of information upon the digital object at hand (and sometimes completely indeterminable provenance, unlike with books and other atom-based materials), fresh approaches to these new expressions of the same old problems have been necessitated. Fortunately for this author, both of the metadata schema investigated are markedly flexible in their approaches and were able to accommodate most instances of the issues encountered.

Works Consulted

A Metadata Kernel for Electronic Permanence. John A. Kunze. 2001. Accessed February 11, 2005 at <http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Kunze/kunze-final.pdf>

The ARK Persistent Identifier Scheme. John A. Kunze, et al. 2004. Accessed February 11, 2005 at <http://www.cdlib.org/inside/diglib/ark/arkspec.pdf>

Using Dublin Core. Diane Hillman. 2003. Accessed February 11, 2005 at <http://dublincore.org/documents/usageguide/>